DOCUMENT RESUME

ED 091 415 88 TM 003 627

AUTHOR Sweigert, Ray L., Jr.; Schabacker, William H.
TITLE The Delphi Technique: How Well Does It Work in Setting Educational Goals.
SPONS AGENCY Georgia State Board of Education, Atlanta.
PUB DATE [74]
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Community Leaders; *Decision Making; Development; Educational Assessment; Educational Needs; *Educational Objectives; Educational Planning; Feedback; Futures (of Society); High School Students; *Methodology; Participation; Prediction; Reliability; Teachers
IDENTIFIERS *Delphi Technique

ABSTRACT

Starting with a set of existing goals, use of the Delphi technique to establish educational goals in this investigation assumed: convergence in perception over rounds, greater convergence on the second round than on later rounds, the reliability of ranking of goals produced, the necessity of at least three rounds, and the desirability of feedback of own response to participants. In testing these assumptions, three studies were conducted, one with 275 community leaders, one with 429 educators, and one with 369 high school students. The first three assumptions were confirmed. The last two assumptions were shown to be questionable. (Author)

ED 091415

TM 003 627

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DELPHI TECHNIQUE: HOW WELL DOES IT WORK IN SETTING EDUCATIONAL GOALS

Ray L. Sweigert, Jr.
Atlanta Public Schools

William H. Schabacker
Educational Testing Service

A paper presented at the Annual Meeting of the American Educational Research Association, Division H Paper Reading Session, "Approaches to Formulation of Consensus on Educational Goals in Needs Assessment," Chicago, April, 1974.

The research reported herein was conducted pursuant to a grant from the Georgia Board of Education to the Atlanta Board of Education under the provisions of Title III of the Elementary and Secondary Education Act, 1965, as amended.

THE DELPHI TECHNIQUE: HOW WELL DOES IT WORK IN SETTING EDUCATIONAL GOALS?

Ray L. Sweigert, Jr.
Atlanta Public Schools

William H. Schabacker*
Educational Testing Service

Objectives. Objectives were: (1) to establish goals for education in Atlanta, using the Delphi technique; and (2) to determine how well the Delphi technique works in establishing educational goals under the conditions of the Project.

The first objective was a major undertaking of the Atlanta Assessment Project (AAP). AAP is a three-year endeavor to develop techniques and tools for measuring the progress of Atlanta's graduating high school seniors — and those young people in the Atlanta area who are old enough to graduate, but will not — toward the achievement of educational goals relevant to living in the Atlanta of 1985 and thereafter. Administered and operated within the Atlanta Public Schools, the Project is funded under Title III, ESEA. The second objective above was a subordinate undertaking of the AAP. It is this second objective, however, that is the primary focus of this paper, but only in the light of how well the first objective was accomplished.

Theoretical Framework. It is generally recognized that there are two types of forecasting involved in establishing educational goals. One type forecasts what conditions probably will be at a given time in the

* Mr. Schabacker was a consultant with the Atlanta Assessment Project at the time this study was conducted, prior to joining the staff of Educational Testing Service.

future, and the other forecasts what educational goals should be in the light of these probable future conditions (e.g., Rosove, 1968; Weaver, 1971). Both types of forecasting were involved in establishing goals for education in Atlanta, 1985. The first type of forecasting was accomplished through tapping the perceptions of experts through position papers they had written about the future in Georgia in their respective fields. The second type of forecasting was accomplished through use of the Delphi technique. There is precedent for the use of the Delphi technique in forecasting what educational goals should be (e.g., see Cyphert and Gant, 1970; and Uhl, 1971).

The Delphi technique was developed by Rand Corporation for use in answering questions about the future when a great deal of uncertainty and complexity surround the area of concern (Dalkey, 1970). The procedure calls for iteration in eliciting perceptions from participants, so that they make a series of judgments, each successive one being made in the light of a summary of judgments of all participants on the previous round. This process is designed to produce increasing accuracy of judgment and increasing agreement among participants from round to round. Rosove (1968), in evaluating 21 different techniques for predicting the future, concluded that the Delphi technique was among the five potentially most useful methods of forecasting that might be applied to the functions of a center for educational policy research. The other four methods require more information and more certainty about the future than the Delphi technique does. Parenthetically, it may be noted that the study of educational goals is a critical function of educational policy research.

In the present study, data were collected to determine the extent to which a number of basic assumptions behind the use of the Delphi technique

were operative. These assumptions may be stated in the following terms:

(1) The process of making successive judgments with feedback produces convergence in perception among the members of a Delphi panel. This proposition is perhaps the most basic of all to the Delphi technique and is widely recognized (e.g., Dalkey, 1970; Uhl, 1971).

(2) Convergence on the second round of judgments (the first round with feedback) is greater than on subsequent rounds. This phenomenon was reported by Cyphert and Gant (1970) and by Uhl (1971).

(3) A reliable ranking of goals may be generated through use of the process. The reliability of Delphi judgments has been reported by Dalkey (1970) and by Uhl (1971). In the present study, emphasis was on the reliability of a ranking of goals based on the judgments, rather than on the reliability of the judgments themselves. The concern was the relative importance of a given goal, not its particular value on a scale of importance.

(4) It is necessary to conduct three or more rounds of a study to produce reliable, convergent results. Since the Delphi technique is based on iteration, the question of how many rounds is necessary, or desirable, has significance. If the process begins with respondents generating the initial goal statements themselves, then one more round will be required, naturally, than if the respondents are presented with a structured questionnaire on the first round. The question revolves around the number of times that feedback preceding judgment is required in order to obtain a satisfactory result. Cyphert and Gant (1970) have raised the question whether or not more than one judgment with feedback is needed.

(5) Participants should be provided with feedback of their own individual last responses, as well as of the last responses of the group as a whole, to facilitate their judgments. Though the usual Delphi procedure

includes feedback of participants' own last responses to them, Uhl (1971) reported an elimination of this aspect of feedback from the method. Uhl commented that emphasizing a participant's previous response, especially if it tended to differ from that of the group as a whole, could make some participants defensive. It was hypothesized in this study that feedback of participants' own individual last responses would tend to reduce the convergence of perceptions about goals.

Data Source. Three studies were conducted using the Delphi technique. One involved professional, technical, managerial, and community leaders in the Atlanta area. The occupational divisions at the professional, technical, and managerial levels presented in the Dictionary of Occupational Titles (1965) were used heuristically for structuring the selection of respondents. Several other categories of respondents were added to provide for individuals in public service roles that were primarily political in nature, e.g., members of the Atlanta Board of Education, members of the Atlanta Board of Aldermen, and state legislators from the Atlanta area. Of the approximately 400 persons invited to participate in this study, 275 completed all three rounds.

The second study involved high school teachers, counselors, principals, and other administrators directly involved with instruction in the Atlanta Public Schools. Teachers were selected to be representative of the entire range of subject matter in each of the 25 high schools then in the Atlanta system and also representative of the racial and sexual distribution of teachers within each high school. All principals and other administrators that were directly involved with instruction were asked to participate. Of the 445 that were invited to take part in the second study, 429 completed all three rounds.

The third study involved high school student leaders selected to represent the 25 high schools and the distribution of students by race and sex within each individual school. Of the 375 students invited to participate, 369 completed all three rounds.

The Delphi technique has usually been employed with relatively small groups of participants. However, Cyphert and Gant (1970) and Uhl (1971) report using much larger groups, 400 in the former study and almost 1,000 in the latter. In the three studies reported here, a total of 1,073 respondents completed all three rounds.

It has usually been the case that groups of experts have been impaneled as participants in a Delphi study. In both of the studies just cited, however, the expertise of respondents was de-emphasized. The results of investigation by Brown, Cochran, and Dalkey (1969), as reported by Uhl (1971), in which students were used as participants, suggest that nothing of significance is lost by including less knowledgeable persons as long as there are some participants who are knowledgeable.

Perhaps expertise is not a critical criterion for selection of a respondent in a study that is concerned with what should be. Perhaps a more important question than who is expert is what kinds of persons should be involved in deciding public policy. The question is as much political as technical, if not more so. Discussions of the accuracy of judgment (see Weaver, 1971) seem less applicable to the question of what should be than to the question of what may be.

Be that as it may, the three groups of respondents included in the Delphi studies in the Atlanta Assessment Project were perceived to have special areas of expertise related to education. It was felt that among the professional, technical, managerial, and community leaders of Atlanta resides the competence to make judgments about the relative importance of specific educational goals in the light of probable future conditions in the Atlanta area. It was thought that probably no group was any more

qualified than this one to make such judgments. Among the teachers and administrators of the Atlanta Public Schools resides another kind of expertise, an understanding of the educational system and what it can do, and an understanding of students. Among the students resides a still different kind of expertise, for they are the ones who are living and experiencing the learning process. The student himself has perceptions of educational goals that, if for no other reason than his unique perspective as a learner, should be included in a Delphi study of educational goals.

Methods and Techniques. The starting point in establishing educational goals for the Atlanta of 1985 was a set of 86 previously identified goals that had been adopted for the State as a whole by the Georgia Board of Education (Advisory Commission on Education Goals, 1970). These goals had been derived from position papers written about probable future conditions in the State by experts in a number of areas of concern (Schabacker et al, 1970). A questionnaire designed to elicit a judgment about each of the 86 goals on a six-interval scale of importance was presented to each participant on three successive rounds. Importance was considered in terms of preparing young people to live in the Atlanta of the future. In the first study, involving professional, technical, managerial, and community leaders, each respondent was interviewed personally once a week for three weeks. In the study involving students, the questionnaire was group-administered every two weeks over the three rounds. In the educator study, the questionnaire was handled in a variety of ways, all documented, from group administration to participant self-administration. What participants did in each of the three rounds in evaluating goals is described below:

Round One: To establish a future-oriented frame of reference in making judgments about the relative importance of goals, each participant was asked to read a short essay containing abstracts of the position papers that were used in the derivation of the goals. In responding to the questionnaire, each participant judged the relative importance of each of the goals in the questionnaire and then wrote down any additional goals that he felt were very important and should be included.

Round Two: Each participant was given an opportunity to again read the essay containing the abstracts of the position papers about the future of Georgia if he so desired. Each participant responded to the same questionnaire as in the first round, but with a difference. For each goal, the response category that was selected by the most participants in the first round — the modal response — was encircled. Participants were asked to write in a "comments column" in the questionnaire their reasons for judging any particular goal to be either more important or less important than the modal response. Additional goals suggested in Round 1 were submitted to participants in an additional goals questionnaire that required judgments on the same scale of importance as that used with the initial 86 goals.

Round Three: Each participant was again given an opportunity to review the essay containing the abstracts of position papers about the future of Georgia if he so desired. The questionnaire used in the third round was the same as that used in the first two rounds, with appropriate response categories encircled to indicate the modal responses made in the second round. To further aid participants in making their final judgments, a summary of comments about each goal was presented with the questionnaire.

This summary contained reasons given in Round 2 for judging each goal to be more important or less important than the modal response. An additional goals questionnaire was administered in Round 3 also.

In the first study, involving professional, technical, managerial, and community leaders, a further dimension of the design was differential feedback of participants' own individual last responses. In order to obtain data on the effects of providing a participant with his own last response to each goal, as well as the modal response for the entire group, four variations were employed, as shown in Table 1. One group of participants received their own last responses to each goal in both the second and the third rounds. A second group of participants received their own last responses in the second round, but not in the third. A third group received their own last responses in the third round, but not in the second. A fourth group received their own last responses in neither the second nor the third rounds. A hypothetical rank ordering of groups as to expected degree of convergence was developed. Members of the panel were randomly assigned to these four treatments.

Results and Conclusions. Analysis of data depended heavily upon non-parametric methods. For a general discussion of the techniques employed here, see Siegel (1956).

The initial set of goals and the set of additional goals suggested by participants were each rank ordered on the basis of the mean importance of each goal as seen by community leaders, by educators, and by students respectively. An overall ranking within each of the two sets of goals was computed by taking the mean of the mean importance ratings across the three groups for each goal and then ranking these.

In ranking 86 goals on the basis of mean importance registered on a six-interval scale, the reliability of the ranking is a fundamental question (Assumption 3 in the presentation of theoretical framework). To determine reliability, each of the three major groups of participants — community leaders, educators, and students — was randomly divided into halves; and the goals were ranked separately for each half. The Spearman rank correlation technique was employed to determine the correlation in ranking between the halves of each group of participants. The resulting coefficients, computed for all three rounds, ranged from .96 to .99, as can be seen in Table 2.

To test the convergence assumption (Assumption 1), the Wilcoxon matched-pairs signed-ranks test was used to determine whether the standard deviation of the judgments about goals became smaller from Round 1 to Round 2 to Round 3. It was found that the signed difference between the standard deviation of Round 1 judgments and that of Round 2 judgments for each goal in the initial set of 86 goals was positive in every case. Thus, unequivocally, convergence did occur in Round 2. Using the same approach, it was found that the signed difference between the standard deviation of Round 2 judgments and that of Round 3 judgments for each goal was positive in 82 of the 86 cases, leaving no doubt that convergence occurred in Round 3 also.

To test the assumption that convergence in Round 2 is generally greater than in Round 3 (Assumption 2), the Wilcoxon matched-pairs signed-ranks test was again used, this time to determine whether the signed difference between the standard deviation in Round 1 and that in Round 2 was generally greater than the signed difference between the standard deviation in Round 2 and that in Round 3 across goals. It was found that $S.D._1 - S.D._2$, where the subscripts indicate the number of the round, was greater than $S.D._2 - S.D._3$

for 83 of the 86 goals, providing very strong evidence that the assumption was correct.

A major concern in considering the question of how many rounds are necessary, or desirable, in a Delphi study (see Assumption 4) is the degree of correlation between the results of the different rounds. Spearman rank correlation coefficients were computed to determine the correlation in the overall ranking of goals from one round to the next. It was found that the Spearman rho for Rounds 1 and 2 was .98, for Rounds 2 and 3 was .99, and for Rounds 1 and 3 was .98. If ranking is the major concern, one round may be enough.

In examining the data for the effects of differential feedback of own response (see Assumption 5), a Kendall coefficient of concordance was computed for each of the three rounds to determine the extent of agreement in the ranking of goals among the four groups receiving different patterns of feedback. These coefficients were .95, .96 and .96 respectively, showing a very high level of agreement among the four groups.

Perhaps the most interesting question to be answered had to do with whether or not there was differential convergence of perception among the four groups. It was hypothesized that reminding a participant of his own last response at the same time that he received the modal response for the entire group would tend to reduce the likelihood that he would select the modal response as his judgment of the goal in that round. Applying this proposition to each of the four treatment groups shown in Table 1 led to the conclusion that convergence would be least for those who received their own individual last responses in both the second and third rounds, and greatest for those who did not receive their own last responses in either of these rounds. Since convergence in the second round, or the first round with feedback, tends to be greater than that in the third round, it was further proposed: participants who received their own last responses in the second round (inhibiting convergence when it would otherwise tend to be

relatively large), but not in the third, would tend to converge less than those who received their own last responses in the third round, but not in the second. This line of reasoning produced a hypothetical rank ordering of the four groups as to expected degree of convergence.

To test this rank ordering of groups, the signed difference in standard deviation between Rounds 1 and 2, Rounds 2 and 3, and between Rounds 1 and 3 were computed for each group in respect to each goal. The Kruskal-Wallis one-way analysis of variance was used to determine whether or not there were variations among the four groups in each of these sets of differences in standard deviation. The results can be seen in Table 3.

The first row of the table shows the relative degree of dispersion in perception in each group in Round 1 when all participants, regardless of group, received exactly the same stimulus. The sums of ranks based on the standard deviations in Round 1 are presented here only for purposes of comparison. It should be kept in mind that for each row of the table the indicated measures for all groups on all goals were put into a single rank order.

Reference to Table 1 shows that in Round 2, Groups A and B should be similar in having relatively little convergence, since both groups received their own individual last responses, and Groups C and D should be similar in having relatively more convergence, since both of these groups did not receive their own individual last responses. The pattern of sums of ranks in the second row of Table 3 shows this tendency clearly. Groups A and B exhibited some convergence, but less than did Groups C and D. The Kruskal-Wallis test for this row produced a very large H of 199.97, showing that the differences among sums of ranks in the row are highly significant

statistically. With 3 degrees of freedom, it requires an H of 16.27 to be statistically significant at the .001 level.

Reference once more to Table 1 shows that in Round 3 the difference in convergence between Groups A and D should again be large, but that Group B should show relatively more convergence and Group C relatively less than in Round 2. The pattern of sums of ranks in the third row of Table 3 shows that Group B had the greatest convergence of all groups in Round 3 and that Group C had the least, thus confirming part of the expectation. However, Group D showed an unexpected reduction in relative convergence, whereas Group A showed more convergence than anticipated. The Kruskal-Wallis test of differences among sums in this row again produced a very large H (119.93).

The fourth row of Table 3 shows the overall effects of differentiated feedback of own last responses from Round 1 to Round 3. It can be seen that Group A had the least convergence, with Group B second in this respect and Group C third, as expected. Group D was the only one that did not conform to expectation. The Kruskal-Wallis test produced an H of 92.62 for the fourth row.

Discussion. The hypothesis that feedback of participants' own last responses tends to reduce the convergence of perceptions was confirmed. However, the expected rank ordering of treatment groups in respect to overall degree of convergence across rounds had one group out of its hypothetical place. Group D conformed to expectation in Round 2, but not in Round 3. The overall degree of convergence for Group D was less than that of Group C. It is interesting also that Group A had relatively more convergence in Round 3 than in Round 2, in fact more in that round than did Group D.

It would appear that there was a tendency for a stimulus situation to be somewhat less effective the second time around. Groups A and D were the two groups that respectively experienced the same simulus situation in both Rounds 2 and 3. Perhaps Group D converged less than anticipated in Round 3 because, having converged appreciably in Round 2, there was not as much room left to converge as there had been previously. Further, having experienced pressure to change in Round 2 through confrontation by modal responses, and having tended to yield to it, Group D may have been likely to develop some resistance to further change.

This explanation is supported by the fact that 39 of the 86 difference scores used in determining the third row sum of ranks for this group were negative; that is, for almost half of the goals, dispersion was greater in the third round than in the second, showing some movement away from modal responses. This was true in the third round for Group C also, for which 45 of the 86 difference scores were negative.

Group A, however, having experienced the likely discomfort of discrepancies between modal responses and own responses on the second round, and being faced with such discrepancies again on the third round, may have found that commitment to own response tended to be more difficult to maintain the second time around and that movement toward the modal response was a somewhat more satisfactory resolution of the situation than it had been before. The discussion is very speculative at this point, of course, and does not really explain anything about the behavior of participants in Groups A and D in the third round. However, it does offer some possible mechanisms that may have been at work.

Summary. Starting with a set of previously identified goal statements, use of the Delphi technique in establishing educational goals in this investigation assumed convergence in perception over rounds; greater convergence in the second round than in later rounds; the reliability of a ranking of goals produced; the necessity of at least three rounds; and the desirability of feedback of own responses to participants. In testing these assumptions, three studies were conducted, one with 275 community leaders, one with 429 educators, and one with 369 high school students. The first three assumptions were confirmed. The last two assumptions were shown to be questionable.

Importance of the Study. The importance of goal-setting in education today is widely recognized. Different methods of goal-setting are being tried with varying degrees of success. The Delphi technique is one of these methods. The findings of this study shed some light on the usefulness of the method. If a study starts from a set of previously identified goal statements and Ns of good size, the finding that there was a high correlation between the ranking of goals in Round 1 and that in Round 3 suggests that one round may be sufficient. If it is considered desirable to produce convergence, however, then a three-round study would be in order. The finding that feedback of own last responses reduces convergence leads to the conclusion that feedback of own last responses should not be included in the design if convergence is desired. Given Ns of good size, a reliable ranking of a large number of goals can be obtained using the methods reported here.

TABLE 1

DIFFERENTIAL FEEDBACK OF OWN RESPONSES
TO MEMBERS OF THE DELPHI PANEL

	Received Own Responses From Previous Round	
Group	Round 2	Round 3
A	Yes	Yes
B	Yes	No
C	No	Yes
D	No	No

TABLE 2

RELIABILITY OF THE RANKINGS OF INITIAL GOALS

Group	Spearman Rank Correlation Coefficients		
	Round 1	Round 2	Round 3
Community Leaders	.97	.98	.99
Educators	.99	.98	.96
Students	.97	.98	.98

NOTE: In determining the reliability of the rankings, each group of participants was randomly divided into halves, and a ranking of goals was developed for each half. The correlation between the rankings for the halves was then computed for each group.

TABLE 3

CONVERGENCE OF PERCEPTION AS A FUNCTION OF DIFFERENTIAL FEEDBACK OF OWN RESPONSE, SHOWING SUMS OF RANKS IN FOUR KRUSKAL-WALLIS ONE-WAY ANALYSES OF VARIANCE

	Sums of Ranks Across Goals			
Ranked Measure	Group A	Group B	Group C	Group D
$S.D._1$	16,238.0	15,762.5	14,774.5	12,565.0
$S.D._1 - S.D._2$	<u>9,215.5</u>	<u>7,958.0</u>	23,645.5	18,521.0
$S.D._2 - S.D._3$	<u>15,483.5</u>	22,531.0	<u>8,720.5</u>	12,605.0
$S.D._1 - S.D._3$	8,222.5	14,573.0	20,655.0	15,889.5

NOTE: The subscripts in the lefthand column indicate the round of the Delphi study. A Kruskal-Wallis test was run on each row of the table. For each test, the smallest difference in S.D. was given a rank of 1, and the largest was given a rank of 4 x 86, or 344 (across 86 goals for 4 groups). Therefore, the larger a sum of ranks in the second, third, and fourth rows, the greater the relative convergence of that particular group in the indicated round. The underlining of a sum of ranks signifies feedback of own response for a given group on the round in question.

References

Brown, B., Cochran, S., and Dalkey, N. The Delphi Method, II: Structure of Experiments. The Rand Corporation, RM-5957-PR, June, 1969.

Cyphert, F. R., and Gant, W. L. "The Delphi Technique: A Tool for Collecting Opinions in Teacher Education." The Journal of Teacher Education, 21(3), 1970, 417-425.

Dalkey, N. "Use of the Delphi Technique in Educational Planning." Herald, (Newsletter of the Educational Resources Agency, Sacramento, California), 4(2), 1970.

Dictionary of Occupational Titles, Vol. II, Occupational Classification, (third edition). U. S. Department of Labor, 1965.

Georgia Advisory Commission on Educational Goals of the State Board of Education. Goals for Education in Georgia. Final report to the State Board of Education. Atlanta: Georgia Department of Education, 1970.

Rosove, P. E. "A Provisional Survey and Evaluation of the Current Forecasting State of the Art for Possible Contributions to Long-Range Educational Policy Making." Appendix A., A Pilot Center for Educational Policy Research. Final report to the U. S. Office of Education. Santa Monica, Cal.: System Development Corp., 1968.

Schabacker, W. H., et al. (Editors). Focus on the Future of Georgia, 1970-1985. Papers prepared for use by the Advisory Commission on Educational Goals of the State Board of Education. Atlanta: Georgia Department of Education, 1970.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Co., Inc., 1956.

Uhl, N. P. Encouraging Convergence of Opinion, Through the Use of the Delphi Technique, in the Process of Identifying an Institution's Goals. Durham, N. C.: Educational Testing Service, Southeastern Office, 1971.

Weaver, W. T. "The Delphi Forcasting Method." Phi Delta Kappen, 52(5), 1971, 267-271.